DOCUMENT RESUME

ED 291 771 TM 011 056

AUTHOR Webber, Charles F.
TITLE Program Evaluation: A Review and Synthesis.
PUB DATE 87
NOTE 28p.
PUB TYPE Information Analyses (070)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Educational Assessment; Ethics; Evaluation Utilization; Holistic Evaluation; Needs Assessment; *Program Evaluation; Qualitative Research; Statistical Analysis
IDENTIFIERS Stakeholder Evaluation

ABSTRACT

This paper reviews models of program evaluation. Major topics and issues found in the evaluation literature include quantitative versus qualitative approaches, identification and involvement of stakeholders, formulation of research questions, collection of data, analysis and interpretation of data, reporting of results, evaluation utilization, and ethical issues in program evaluation. There appears to be a trend toward the synthesis of program evaluation models. This focus incorporates concern for rigorous design, a combining of quantitative and qualitative methodologies, respect for the perspectives of all participants, pragmatism, active stakeholder involvement, and social value. The features of this focus proceed through the following steps: (1) study of program content; (2) establishment of stakeholder commitment and involvement; (3) focusing of the evaluation; (4) formulation of an evaluative design; (5) data collection; (6) holistic and statistical analyses; (7) interpretation of results; and (8) program modification. The process is circular and continuous. (TJH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED291771

PROGRAM EVALUATION:

A REVIEW AND SYNTHESIS

by

Charles F. Webber, Ph.D.

1987

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Charles Webber

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

TM011 056

BEST COPY AVAILABLE

Different models of program evaluation have been developed for a variety of functions. Goal-oriented evaluation, for example, fits situations in which participants wish to assess student progress and monitor effectiveness of particular innovations. Decision-oriented evaluation focuses on informed decision making. Transactional evaluation is organized around program processes and the value positions held by key participants. Efforts to explain educational effects and devise instructional strategies can be categorized as evaluation research. Goal-free evaluation represents efforts to assess program effects without being limited by a program's conceptual framework. Adversary evaluation offers competing stakeholders access to the same program information (Cronbach, 1982; Morris & Fitz-Gibbon, 1978).

Program evaluations are also commissioned for more general reasons. Stakeholder uncertainty and confusion can be allayed. Decisions can be made in a more timely, systematic and informed manner. Evaluators can help clients avoid planning based on what is fashionable. Understanding of educational diversity and complexity can evolve. Evaluations offer interest groups a common language, i.e. terms, data models, and orientations, that add coherence to discussions. Program evaluation helps an organization refocus when there is a danger of a program, per se, supplanting client need as a raison d'etre. Programs can be rendered more efficient, productive, and effective. Program credibility and acceptance can be augmented through the gathering of supportive evidence (Cronbach, 1982; Bennet & Lumsdaine, 1975; Sarason, 1982; Stake, 1967; Weiss & Bucuvalas, 1977).

One justification offered for program evaluation is accountability. This can be interpreted as teacher, student or program accountability. Other constructs associated with accountability are productivity, cost-effectiveness, quality control, and improved standards. These concepts can

be useful and relevant to policy shapers. However, accountability can imply "that everybody, regardless of qualifications, has the right to pass judgement on the teacher's performance" (Bowers, 1974, p. 138). More positively, taken-for-granted beliefs underlying programs can be critically examined, new issues can surface, thinking can be revised, and policies can be redefined. Program participants can be encouraged to assume responsibility for their own situations (Bowers, 1974; House, 1980; Sarason, 1982; Weiss & Bucuvalas, 1977).

Program contextual variance is nearly limitless. Site adaptation or the "mutation phenomenon" (Berman & McLaughlin, 1974, p. 10) means that any one evaluation will encounter a plethora of influences. Therefore, features from a variety of models have been incorporated into the following discussion, which has been influenced by the works of L.J. Cronbach, E.R. House, E.G. Guba, Y.S. Lincoln, and M.D. Patton.

Quantitative Versus Qualitative Evaluation Research

The traditional view of evaluation research concentrates on results and follows three steps:

"1. Two or more conditions are in place, at least one of them being the consequence of deliberate intervention.

2. Persons or institutions are assigned to conditions in a way that creates equivalent groups.

3. All participants are assigned on the same outcome measure(s)" (Cronbach, 1982, p. 24).

One example of this view is the systems analysis approach, which directs results toward managers and economists. It assumes program goals are agreed upon by all stakeholders, that cause-effect linkages can be established, and that outcome variables can be quantified; efficiency is its goal (House,

1980).

Patton (1978) compares traditional evaluation research to the natural science format of hypothetico-deductive methodology. This dominant mode characterizes good research as incorporating quantitative measurement, experimental design, and multivariate, parametric statistical analysis. These characteristics have been transposed from the basic statistical and experimental techniques of agricultural experimentations (Partlett & Hamilton, 1976; Patton, 1978).

Evaluation research has moved away from the view that the only worthwhile study is one which yields reliable, quantifiable data (Cronbach, 1982; Cronbach, et. al., 1980; Guba & Lincoln, 1981; House, 1980; Patton, 1978, 1980). Qualitative research has become more common and increasingly accepted. Purportedly, evaluators now understand that people formulate plans, values, and purposes which are influenced by emotions, cultures, and life experiences.

Evaluators with an anthropological bent believe modern, pluralistic societies must be examined through in-depth, open-ended interviews and personal observations which yield qualitative data that can be analyzed holistically. They say this type of inquiry leads to understanding of particular situations as opposed to predictive validity. Qualitative evaluators insist that site variation negates the applicability of quantitative results to other settings. Instead, evaluations should take advantage of local conditions and serendipity in their efforts to reach understanding (Ager, 1980; Geertz, 1973; Guba & Lincoln, 1981; House, 1980; Patton, 1978; Pelto & Pelto, 1978; Powdermaker, 1966).

Case studies exemplify the utility of qualitative studies because they are politically sensitive and "more likely to be attended to than are typical

evaluation reports" (Guba & Lincoln, 1981, p. 378). Case studies can focus on diverse program aspects and lead to community learning. Stake (1977) says this makes the investigator a teacher who helps refine clients' perceptions of a program, instead of trying to establish and maintain evaluator power and expertise. Thus, the evaluator-teacher comes to value, just as ethnographers do, subjective data garnered through observation and interpretation of participant behaviors (Cronbach, 1982).

Although he highly values the case study approach, House (1980, p. 247) cautions that it is "no panacea and entails a distinctive set of problems of its own." Case study theory and methodology should be carefully examined (House, 1980; Patton, 1978). With these cautionary notes in mind, Guba and Lincoln (1981, p. 377) extol the freedom of case studies which allows situational "vibes" to be picked up and explored. This, they feel, is not as possible in tightly controlled studies which insist that everything be scientifically documented; that is, the "case study provides a vehicle for the transference of that kind of wordless knowledge" (Guba & Lincoln, 1981, p. 377).

Guba and Lincoln (1981) have three suggestions for qualitative evaluators. First, case study records should be kept current and clearly show defensible links between raw data and conclusions. Second, interviews should be carefully documented to show they were reliably and validly conducted. Third, instead of apologizing for the subjective nature of case studies, evaluators should list the advantages inherent in the subjectivity:

- Questions can be restated if not initially understood.
- Interviews are personal.
- Sensitive topics can be dealt with compassionately.
- The affective responses of informants can be noted.

- Nonverbal clues can be studies.

- Rich contextual information can be gathered.

There are similarities between the experimental and naturalistic evaluation camps. While experimental evaluators give more credence to quantified data than the naturalistic evaluators would, they both believe that society should attempt to progress. Further, they both acknowledge that change can be well intentioned but harmful. Technical similarities include a shared concern for sampling, question formulation, and quality. Evaluations should cross freely between the two categories because one is not better than the other; choice should be made on the basis of suitability to the program under study (Cronbach, 1982; Cronbach, et. al., 1980; Bennett & Lumsdaine, 1975; Kuhn, 1970; Patton, 1980). As Patton (1978, p. 235) says, "There is no single factor or set of factors that can solve the mystery of human behavior, no one answer to the fundamental philosophical question: why do people do what they do? (Nor is there a single answer to that most fundamental of governmental questions: how do we get people to do what we want them to do?)"

There have been calls for the use of multiple data gathering methods (Campbell & Fiske, 1959; Cronbach, 1982; Cronbach, et. al., 1980; Denzin, 1978; Guba & Lincoln, 1981; Patton, 1980). Triangulation, as this is called, allows different aspects of a program to surface. Cronbach, et. al. (1980, p. 222) state, "Those who advocate an evaluation plan devoid of one kind of information or the other carry the burden of justifying such exclusion."

Identifying and Involving Stakeholders

Significant program studies will produce results that will increase or decrease the power of individual interest groups. Evaluation researchers can anticipate that individuals whose power is increased will support and

defend the study and vice versa (Guba & Lincoln, 1981). An evaluation should be undertaken only when the political system indicates it will give serious consideration to results generated by the study. The political system may include groups as diverse as voters, managers, operating personnel, and policy makers, which means considerable effort must be exerted by the evaluator to ascertain levels of audience receptivity (Cronbach, 1982; Guba & Lincoln, 1981).

Cronbach (1982) says the degree to which an evaluation is successful is the extent to which the interest groups are able to resolve conflicts intelligently. This may seem contrary to Guba and Lincoln's (1981, p. 299) statement that, "Evaluation is always disruptive of the prevailing political balance." However, this becomes clear when the level of political dissent is seen as a question of degree. In other words, a successful program evaluation will reduce the level of stakeholder disharmony over a particular issue, though it cannot expect to satisfy everyone involved.

If an evaluation is to reduce disharmony, then the evaluator should involve stakeholders in the identification of contentious issues (Cronbach, 1982; Guba & Lincoln, 1981; Patton, 1978). Further, the evaluator should try to become well acquainted with stakeholders' beliefs and values because "different people have different appetites for different information" (Stake, 1973, p. 304).

Guba and Lincoln (1981, p. 308) suggest three things to consider during interest group identification:

"1. Who are the presumed direct beneficiaries of the evaluand?

2. Who are the indirect beneficiaries?

3. What groups might, as a result of the evaluation, be persuaded to adopt or adapt the evaluand in their own settings?"

Considering negative effects will also help identify stakeholders.

Failure to involve audiences at the beginning of an evaluation may automatically cause critical questions to be overlooked, study results to be suspect, and methods to be criticized. It may also be unfair to audience members because, "The act of evaluation provides a political legitimation difficult to achieve in other ways" (Guba & Lincoln, 1981, p. 306). Exclusion of crucial audiences may misapply that legitimacy. However, some audiences may have motives for trying to derail an evaluation.

Audience involvement means persons with the power to facilitate or hinder entree (the stage where stakeholder commitment is sought) need to be committed to the study. These strategically located individuals are called "gatekeepers" (Guba & Lincoln, 1981, p. 290). Each gatekeeper will require an explanation of the evaluation before the research can effectively continue.

Gatekeepers do not have to be people of formal position and authority. They should, however, be enthusiastic, committed, competent, interested, and aggressive. Patton (1978, p. 71) suggests that "more may be accomplished by working with a lower level person ... than in working with a passive, disinterested person in a higher position." Failure to work with gatekeepers can mean an evaluation is really not targeted at all, resulting in reduced utilization potential.

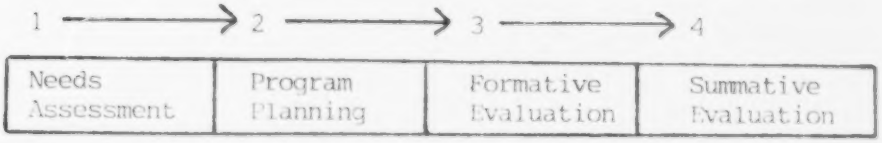
Involving gatekeepers who have a genuine interest in research data is called considering the "personal factor" (Patton, 1978, p. 70; House, 1980, p. 64). Recognizing the personal factor shows an understanding of how decision making is a personal and political process, rather than strictly a scientific and rational process.

Varied audience information needs imply multiple criteria; multiple

methods of data collections; different styles of data analysis, interpretation, and reporting; and varied satisfaction with study conclusions (House, 1980). This wide range of interests and needs stresses the importance of formulating a contract with evaluation audiences (Cronbach, 1982; Guba & Lincoln, 1981). Minimally, contracts should address the following issues, regardless of contract complexity: "identification of the sponsor or client, identification of the entity to be evaluated, purpose of the evaluation, sanction (from relevant parties), audiences, methods of inquiry, emergent design, access to records, confidentiality and anonymity, evaluator autonomy, reporting, and technical specifications" (Guba & Lincoln, 1981, p. 271-282).

Formulating Research Questions

Scriven (1967) delineated two types of evaluation - formative and summative. Formative evaluation tries to collect information for program development and improvement and summative leads to more final judgments about program effectiveness. Patton (1978) suggests Scriven's distinction between formative and summative evaluation may be artificial. Instead, all evaluations can be viewed as formative. That is, evaluation of a program's outcomes can and should be used formatively by asking why the program was effective, thus assisting others who may be considering implementing or improving the same or a similar program. Nevertheless, the formative-summative distinction is widely accepted as depicted in the figure below, developed at the Center for the Study of Evaluation (CSE), University of California, Los Angeles:



Stages of the CSE Evaluation Model (Morris & Fitz-Gibbon, 1978, p. 8)

The formative-summative distinction is important when evaluators formulate evaluation questions. To do this they must consider evaluation purposes, information uses, types and amounts of evaluation data likely to be generated, and alternative actions that will be open to decision makers once the evaluation is complete. This stage is followed by one in which evaluators decide whether appropriate data can be gathered to answer a question, the degree to which the research questions predetermines or suggests answers, how badly decision makers want or need the answers for their own or others' decisions, and how relevant the question is for future action (Cronbach, 1982; Patton, 1978).

Program goals can be a source of questions, especially during early phases of evaluations. However, program goals may be politically decorative, not necessarily reflecting real goals nor hinting at unwanted side effects (Cronbach, 1982; Cronbach, et. al., 1980; Schultze, 1968). If goals are used as a source of questions, then evaluators would do well to look for goals that are unrealistic, particularly well met, or ones program staff did not even try to reach. Cronbach (1982) also warns against setting quantitative expectations for goal attainment. He suggests satisfactory levels should be negotiated after the assessment, when quantitative data can be seen in light of the qualitative data also garnered.

In the initial stage of formulating evaluation questions, many issues will be highlighted. When investigators begin to plan the kinds of observations they need to make, they usually see that far more should be done than can be done. Patton (1978, pp. 80-81) claims, in reference to university based investigations, that, "Professors have trouble getting graduate students to analyze less than the whole of human experience in their dissertations." This general sentiment is echoed by Cronbach, et. al. (1980) who suggest

that one evaluation or even a series of evaluations will not end an argument about a program. Instead of trying to do too much, program evaluators should remember that relevance is important and limit their investigations accordingly.

Those who feel they would like to investigate a large number of issues should consider several things. First, even though efforts will probably be made to give equal attention to all questions, some will necessarily end up getting more. Thus, focusing will occur and better that it be guided than haphazard. Second, investigators, especially neophytes, may become overwhelmed at the task they have assigned themselves and become unable to do a good job on any part. Third, small-scale evaluations can more quickly and easily depict the usefulness of program evaluation; future studies can then expect more support than if large-scale ones had failed. Finally, the audience will not need nor want to know absolutely everything about a program. Evaluators should be cognizant of audience attention span (Borg & Gall, 1979; Cronbach, 1982; Patton, 1978).

One danger of attempting too much is that by the time the study is done the situation may have changed and the results be of little more than passing interest. Ideally, a study will provide accurate and perceptive information to decision makers when it will be more useful (Cronbach, 1982; Marris & Rein, 1973).

The influence study data have on subsequent decisions is called leverage (Cronbach, 1982; Cronbach, et. al., 1980; Patton, 1978). Leverage is critical because rigorous methodology, sophisticated statistical analyses, and large samples are worthless if they yield useless information. Determining leverage involves considering the politics of the issue and the import of the decisions affected by the question. Programs sometimes have so much political backing

that even if evidence of their ineffectiveness is discovered it will have no leverage. Evaluators should think carefully about expending energy on investigations or programs that are firmly ensconced in the political milieu (House, 1973).

Cronbach (1982) offers a type of priority scale for deploying investigative effort. His comments have been transformed into this figure:

	high leverage	low leverage
high prior uncertainty	A. essential to include	B. limited resource investment
low prior uncertainty	C. only low cost information should be gathered	D. investigate only incidentally

Priority scale for evaluation questions

Issues which fall into block A must be addressed if the evaluations' credibility is to be maintained. Block B issues should not be allocated a high resource expenditure unless initial investigations yield data so compelling that the issue gets pushed into block A.. The same can be said for block C. Category D questions include those which would yield information that would be interesting but not useful for decision making.

Flexibility should be built into an evaluation plan so that, as events unfold, new variables can be identified as politically salient and worthy of inclusion. Initially, it should be presumed that variables which participants say are important do have leverage (Cronbach, et. al., 1980).

No matter how carefully potential objections are envisioned and countered, unpalatable findings will be attacked from some perspective. Designing a program evaluation is a political process and if one of the decision makers does not want a questions answered, it will find ways to denigrate the answers

which evolve. This underlines the importance of involving decision makers in the formulation of research questions and of keeping the design as open as possible, even though such tentativeness may be uncomfortable for those used to rigid hypothesis-testing research. When a research question is challenged by part of the policy shaping community, the evaluator must decide if the political imbalance and human uncertainty likely to result if the question is pursued are worth provoking (Cronbach, et. al., 1980; Guba & Lincoln, 1981; Patton, 1978).

Evaluators cannot answer all questions and reduce all uncertainty. All an evaluator can hope to achieve is to shed some new light, add some additional information, and increase certainty in some small way (Cronbach, 1982; Cronbach, et. al., 1980).

Gathering Data

There appears to be no single best plan for a particular study even when factors of time and budget are considered. Thus, formulating study designs can be viewed as learning by both evaluators and stakeholders. This is particularly encouraging for novice evaluators who can utilize their relative freedom of movement among all segments of the policy shaping community to generate a picture of the situation more complete than what other participants could generate (Cronbach, 1982; Cronbach, et. al., 1980).

Involving stakeholders in the planning of data gathering will encourage stakeholder commitment. It will also be possible to anticipate satisfaction and dissatisfaction with potential findings and to incorporate appropriate strategies. Investigators should guard against being reduced to technicians by stakeholders, merely applying sampling measurement and statistical analysis skills (Cronbach, 1982; Guba & Lincoln, 1981; Patton, 1978).

Evaluation researchers must also choose between fidelity (answer reliability) and bandwidth (the number of questions for which answers are offered). While focusing resources on one issue will improve fidelity, it is more often appropriate to strike a balance with bandwidth. The degree of balance will vary with each situation and will be determined cooperatively by the evaluator and the policy shaping community (Cronbach, 1982).

Audiences need to be aware of influences wielded by a program's political environment, decision maker uncertainty, and information availability. They can then recognize that evaluation findings can be useful in particular settings but need to be viewed tentatively when generalized to other settings. They will realize that the actual study is substituting for the ideal study that would supply irrefutable data (Cronbach, 1982; Patton, 1978).

Because there are always more methods of inquiry possible than resources available, the range of choices must be reduced. This narrowing will be helped by developing a thorough understanding of the setting: community characteristics, organizational characteristics, staff peculiarities, etc. Further, the purposes of the study need to be re-examined as does the proposed timeline. These considerations will strongly influence the number of variables examined, plus the size and nature of the sample(s) to be studied (Cronbach, 1982; Patton, 1978).

Patton (1978) and Cronbach (1982) say studies should be sensitive to local conditions, not mechanically objective. The understanding they seek comes from being close to the situation, i.e. subjective, resulting in a personalized evaluation, more legitimate in the eyes of program participants. This can help identify interesting and important program features that would not have been initially noted (Cronbach, 1982; Cook, 1981; Berryman & Glennan,

1980). Cronbach (1982) goes so far as to say that it would be unnatural for programs or experimental treatments to be implemented identically in different sites. Examining a program qualitatively means the influences of participant biases and experiences can be noted, as can degrees and nuances of variation from site to site. This is the "ecological correlation" (Cronbach, 1982, p. 99) between a setting and a program.

Focusing on site variation and adaptation of a program tends to make a study less able to be replicated than an experimental design. Cronbach (1982, p. 293) advises the evaluator to "sample those strata that he thinks will predominate in the future," which is not a strategy used in formal experimental design, although it could be part of a quasi-experimental one. Patton (1980, p. 101) agrees and suggests it is possible for "decision makers and evaluators (to) think through what cases they could learn the most from, and those are the cases that should be selected for study." He goes on to say that the least desirable type of sampling is based on convenience.

Attempting to understand program diversity in different settings means, according to Cronbach (1982), evaluation plans and operations cannot be rigidly fixed. However, doing this means results cannot be generalized beyond the group or situation under scrutiny. To avoid this restriction, an investigator can employ random sampling so the data collected will more likely be representative of the larger population. Sample size would depend on the amount one would wish to generalize beyond N and the amount of error acceptable (Cronbach, 1982; Patton, 1980). It is important to remember that the virtues of strong designs should not prompt anyone to think of them as the only worthwhile design (Cronbach, et. al., 1980).

Analysis, Interpretation, and Reporting of Data

All stakeholders should continue to have input throughout the latter

evaluation stages. Manners require that evaluators, who have consulted with audiences earlier in the process, not ignore stakeholders until the final report. Continuous evaluation feedback will constitute more of a learning process for evaluation participants. Supplying information to only a portion of the policy shaping community provides that portion with power in the form of knowledge control. Evaluators should strive to advise the entire audience by seeing that information is thoroughly disseminated and explained. Investigators cannot hope to remove all doubt or confusion surrounding a program by writing a summary report. Issues are generally too complex for this to happen. Unless audiences are kept abreast of the findings as they are uncovered, they may not accept a final report. Evaluation surprises of this sort may in fact increase uncertainty rather than reduce it (Cronbach, 1982; Cronbach, et. al., 1980; Guba & Lincoln, 1981; Patton, 1978).

Guba and Lincoln (1981) say that the complexity of data reporting may vary from stakeholder to stakeholder, particularly when audience sophistication varies. Efforts should be made to inform and explain so partisans do not interpret findings incorrectly and unwisely. Some portions of a program may be doing well and should be praised before shortcomings are identified. Evaluators should consider audience attention span and limit analyses and data sets that probably will not affect decision making, while still preserving a display of evidence and reasoning. Also, having interest groups help in data interpretation will help air differences, misconceptions, and uncertainties (Cronbach, 1982; Cronbach, et. al., 1980).

Evaluators operating from an experimental or quasi-experimental basis tend to believe that presentation of data to interest groups will ensure utilization. Others lean more to the view that people are also influenced

by their previous experiences, thus viewing data subjectively (Cook & Campbell, 1976; Cronbach, 1982; Cronbach, et. al., 1980; Patton, 1978). Audience understanding can be assisted by relating data to other knowledge areas, like folklore, history, community experiences, and common sense (Cronbach, 1982; Lakatos, 1970; Lindblom & Cohen, 1979; Weiss & Bucuvalas, 1977).

Evaluators can incidentally collect colorful and realistic material that will make interim and final reports more vivid. Anecdotal information can lend immeasurably to credibility and utilization potential of study conclusions. This is especially so when information consumers are relatively ignorant of the analyses usually performed upon data by academics (Cronbach, 1982; Cronbach, et. al., 1980). Patton (1978, p. 234) claims this humanistic touch has "received little more than lipservice in most evaluation research."

House (1980) clearly separates analysis and interpretation. He says analysis is the organization of data, construction of statistical tables, and arrangement for presentation. Interpretation is the act of making judgments about what the data mean. Scriven (1967) suggests the evaluator draw conclusions while Cronbach (1982) states that the evaluator should only present data, not making any definitive policy suggestions. House (1980), though, says both should be done, but with the boundaries of each clearly demarcated. This way, audiences can view the analysis separately without the evaluator's interpretations. Cronbach (1982) says that interpreting any type of data conclusively is unrealistic because of site adaptation. This lends credence to Guba and Lincoln's (1981, p. 381) statement that "if the evaluator and the client interact in producting judgments and recommendations, that is, if the judgments and recommendations are produced through a process of negotiation, then each one can make a proper contribution from a posture of integrity."

How much data does an evaluator need before presenting a final report? The better the description of a program the sounder are the judgments based upon that description. On the other hand, the longer a study takes the more chance there is of the situation changing, thus rendering results irrelevant (Cronbach, 1982; Berryman & Glennan, 1980; Thompson & King, 1981). Developing programs need direction as they unfold, not after the fact. Decisions need to be made without full knowledge of their ramifications. The best evaluators can do is to be aware of their limitations, relate data to general experience and theory, and act on this basis when decision makers need assistance.

Program evaluation has been criticized for concentrating on the negative (Freeman, 1977; House 1980). This may be partially due to program evaluations touching political and organizational nerves (Cronbach, et. al., 1980). Patton (1978) and House (1980) believe that the real question is not whether information is positive or negative but whether it is useful to decision makers.

Evaluation Utilization

The literature contains a range of definitions of evaluation utilization. Weiss and Bucuvalas (1977) suggest that utilization has occurred if the evaluator gathers information that advances the decision making process. This can take the form of dramatic and immediate program changes (Alkin, Daillak, & White, 1979; Brown & Braskamp, 1980; Weiss, 1980). This can even result from the efforts of inexperienced evaluators (Cichon, et. al., 1981). Weiss (1972) claims that the rarity of immediate change has contributed to the notion that evaluations have little impact.

However, effects can be more subtle. Knowledge can be gradually assimilated into clients' understandings of important issues. Larger issues may be raised, future studies may be suggested, additional questions may arise,

group solidarity can be enhanced, long term planning may be implemented, problems can be clarified, expectations can be kept realistic, and perceptions altered (Alkin, et. al., 1979; Cichon, et. al., 1981; Patton, 1978; Weiss & Bucuvalas, 1977; Young & Comtois, 1979).

The kinds of effects that an evaluation has will vary according to client expectations, the quality of the information gathered by the evaluator, the client-evaluator relationship, and degree of stakeholder involvement. It may not be possible to see any external evidence of change because only client perceptions may have been altered. Thus, future decisions are indeed influenced, although the connection to the evaluation may not be obvious. Interestingly, effects do not seem to depend upon the use of a formal evaluation model (Alkin, et. al., 1979).

Evaluation data must compete with other information sources. Friends, colleagues, past experiences, and biases influence the degree to which evaluation data are seen as useful (Alkin, et. al., 1979; Guskin, 1980; Weiss, 1980). For example, information users give more credibility to reports from male evaluators than female evaluators, the use of jargon in reports adds to evaluator credibility, and use of the word "researcher" rather than "evaluator" or "content area specialist" creates an impression of objectivity. Also, the closer the audience is to the decision-making role, the more critical it is of the evaluator (Newman, Brown, & Braskamp, 1980).

Womack (1980) says evaluation reports do not compete successfully with other information sources because they are often prepared for professional and not client use. Weiss (1980, p. 231) says this constitutes "poor linkage between researchers and decision makers."

Knorr (1980) says evaluation data can be used in four ways. First, it

can become a base for future decision making. Second, it can induce immediate action. Third, decision makers can use it to merely create the impression that something is being done. Fourth, it can be selectively used to legitimate policy decisions already made. The first two uses fit the definitions of utilization most often encountered in the literature, but the last two have been described as obstacles to utilization (Cichon, et. al., 1981; Weiss, 1980; Williams & Bank, 1981). Other obstacles to utilization include the inability of decision makers to make their needs explicit; inadequate communication channels between evaluators and their clients; decision maker unwillingness to accept evidence that contradicts personal beliefs; the fluidity of community influences, and a poor match between evaluator reports and audience sophistication (Alkin, et. al., 1979; Weiss, 1972, 1980).

Evaluation utilization is not constrained by social science methodology or the changing nature of public issues. Therefore, there remains a relatively untapped potential for decision maker use of evaluation information (Weiss, 1980).

Evaluators have three roles from which they can choose: teacher, assistant and judge. The most critical of these is the teacher role if evaluation utilization is to be maximized (Cichon, et. al., 1981; Guskin, 1980; Patrick, McCann & Whitney, 1981; Weiss, 1972; Wise, 1980; Young & Comtois, 1979). An evaluator in this role must be able to:

- Initiate and maintain interest group cooperation, commitment and involvement.
- Implement the concept of triangulation.
- Keep client expectations realistic and clear.
- Help clients articulate their needs.

- Report findings in an understandable manner.
- Present data at appropriate times.
- Work closely with client groups.
- Establish and maintain a positive working relationship with clients.
- Be flexible.
- Communicate well with interest groups.
- Facilitate client ownership of evaluation results.
- Recognize study limitations.
- Focus on client needs instead of the interests of professional colleagues.
- Take an active role in getting data utilized.
- Employ appropriate research methodology. (Berman, 1981; Cichon, et. al., 1981; David, 1978; Gifford, 1974; Guskin, 1980; Patrick, et. al., 1981; Weiss, 1972, 1980; Williams & Bank, 1981; Wise, 1980; Young & Comtois, 1979).

The scope of these characteristics underlines the difficulty of fostering utilization of evaluation findings. Utilization will not automatically happen; it requires "ingenuity, resourcefulness, and commitment" (Caplan, 1980, p. 9). Further, the nature of these characteristics has prompted Weiss (1980, p. 245) to warn against "the inappropriate acceptance of the results promoted by the most persuasive or charismatic communicator."

Ethics of Program Evaluation

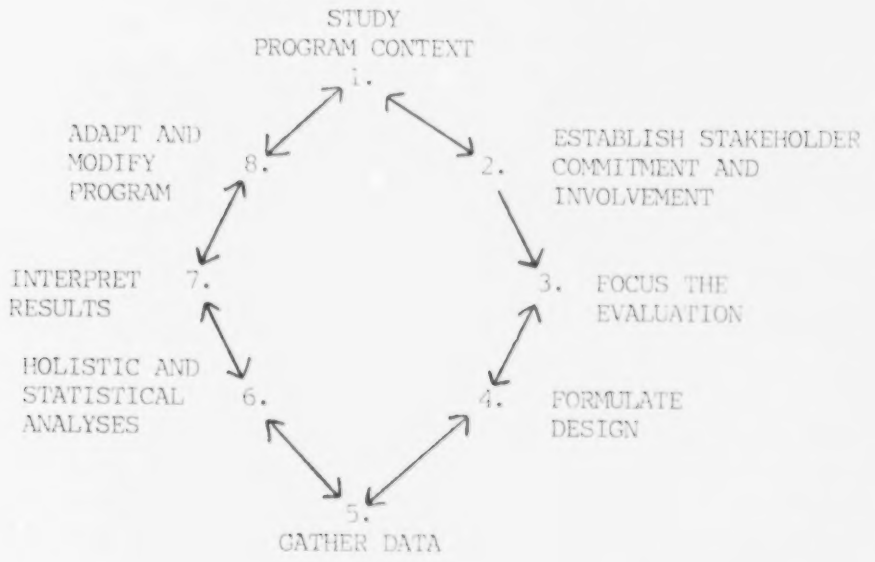
Since leverage and credibility are inextricably linked (Cronbach, 1982), several authors (Care, 1978; Cronbach, et. al., 1980; Guba & Lincoln, 1981; Patton, 1978; Stake, 1977) have sought to delineate guidelines helpful in the maintenance of evaluator credibility. These have been gathered into the following list of professional considerations for program evaluators:

- Stakeholder groups' participation in the evaluation must be of their own volition.
- Stakeholders should be guaranteed their right to have input into the evaluative process.
- Interest groups should be encouraged to honor their commitment to the evaluation that they have expressed through their involvement or renegotiate it through agreed upon channels.
- The agreement or contract should not favor one political entity over another, whether because of evaluator carelessness or the sponsor withholding politically significant information.
- Stakeholders require full and equal access to the accumulation of data and to periodic reports from the evaluator.
- The best interests of the participants should be protected throughout the study.
- Anonymity should be negotiated with subjects before gathering data. When anonymity cannot be guaranteed, subjects should know in advance. They should realize that in studies with small samples it may be possible to identify informants through descriptions or quotations, even though they were not explicitly identified. This can happen despite efforts to combine elements from several cases into a representative case. Evaluators should protect informants by maintaining coded file systems so that individual identities cannot be ascertained by others.
- Evaluators should keep audience expectations realistic, given study constraints.
- Study purposes should be explicitly stated.
- The evaluation must have social value.

A Summary

There is a trend toward synthesis of program evaluation models. This focus incorporates concern for rigorous design, a combining of quantitative and qualitative methodologies, respect for the perspectives of all participants, pragmatism, active stakeholder involvement, and social value.

These features are incorporated into the following process model:



Program Evaluation Process Model

Evaluator-participant interaction is inherent in each stage. The arrows indicate that progress is not automatic. It may be necessary to retrace one or more of the steps before an evaluation can continue. The circular arrangement depicts the continuous nature of evaluation.

REFERENCES

Agar, M.H. (1980). The professional stranger. New York: Academic Press.

Alkin, M.C., Daillak, R., & White, P. (1979). Using evaluations. Beverly Hills: Sage.

Bennett, C.A. & Lumsdaine, A.A. (1975). Social program evaluation: Definitions and issues. In C.A. Bennett & A.A. Lumsdaine (Eds.), Evaluation and experiment: Some critical issues in assessing social programs. New York: Academic Press.

Berman, P. (1981). Educational change: An implementation paradigm. In R. Lehming & M. Kane (Eds.), Improving schools: Using what we know. Beverly Hills: Sage.

Berman, P., & McLaughlin, M.L. (1980). Federal programs supporting educational change (Vol. 2). Santa Monica: Rand Corporation.

Berryman, S.E., & Glennan, T.K., Jr. (1980). An improved strategy of evaluating federal programs in education. In J. Pincus (Ed.), Educational evaluation in the public policy setting. Santa Monica: Rand Corporation.

Borg, W.R. & M.D. Gall. (1979). Educational research. New York: Longman.

Bowers, C.A. (1974). Cultural literacy for freedom. Eugene, OR: Elan.

Brown, R.D., & Braskamp, L.A. (1980). Summary: Common themes and a checklist. In L.A. Braskamp & R.D. Brown (Eds.), New directions for program evaluation. San Francisco: Jossey-Bass.

Campbell, D.T., & Fiske, D.W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. Psychological Bulletin, 56(2), 81-105.

Caplan, N. (1980). What do we know about knowledge utilization? In L.A. Braskamp & R.D. Brown (Eds.), New directions for program evaluation. San Francisco: Jossey-Bass.

Care, N.S. (1978). Participation and policy. Ethics, 88 (4), 316-337.

Cichon, D.J., Callahan, C., & Singh, B. (1981, April). Impact of a process evaluation on an urban school system's policies and practices. Paper presented at the Annual Meeting of the American Educational Research Association (65th), Los Angeles, CA.

Cook, T.D. (1981). Dilemmas in evaluation of social programs. In M.B. Brewer & B.E. Collins (Eds.), Scientific inquiry and the social sciences. San Francisco: Jossey-Bass.

Cook, T.D., & Campbell, D.T. (1976). The design and conduct of quasi-experiments and true experiments in field settings. In M.D. Dunnette (Ed.), Handbook of industrial and organizational psychology. Chicago: Rand McNally.

Cronbach, L.J. (1982). Designing evaluations of educational and social programs. San Francisco: Jossey-Bass.

Cronbach, L.J., Robinson Ambron, S., Dornbusch, S.M., Hess, R.D., Hornik, R.C., Phillips, D.C., Walker, D.F., & Weiner, S.S. (1980). Toward reform of program evaluation. San Francisco: Jossey-Bass.

David, J.L. (1978). Local use of title I evaluations. Washington, D.C: Office of Education (DHEW), Office of Planning, Budgeting, and Evaluation. (ERIC Document Reproduction Service No. ED 187 727).

Denzin, N.K. (1978). The research act. New York: McGraw-Hill.

Freeman, H.E. (1977). The present status of evaluation research. In M. Guttentag (Ed.), Evaluation studies (Vol. 2). Beverly Hills: Sage.

Geertz, C. (1973). The interpretation of cultures. New York: Basic Books.

Gifford, B.R. (1974), December 3). Restructuring the collection, processing and dissemination of educational data: An action plan for change. Brooklyn, NY: New York City Board of Education. (ERIC Document Reproduction Service No. ED 136 829).

Guba, E.G., & Lincoln, Y.S. (1981). Effective evaluation. San Francisco: Jossey-Bass.

Guskin, A.E. (1980). Knowledge utilization and power in university decision making. In L.A. Braskamp & R.D. Brown (Eds.), New directions for program evaluation. San Francisco: Jossey-Bass.

House, E.R. (1973). Epilogue: Can public schools be evaluated? In E.R. House (Ed.), School evaluation: The politics and process. Berkeley, C.A: McCutchan.

House, E.R. (1980). Evaluating with validity. Beverly Hills: Sage.

Knorr, K.D. (1977). Policy makers' use of social science knowledge: Symbolic or instrumental? In C.H. Weiss (Ed.), Using social research in public policy making. Lexington, MA: Lexington Books.

Kuhn, T. (1970). The structure of scientific revolutions. Chicago: University of Chicago Press.

Lakatos, I. (1970). Falsification and the methodology of scientific research programmes. In I. Lakatos & A. Musgrave (Eds.), Criticism and the growth of knowledge. London: Cambridge University Press.

Lindblom, C.E., & Cohen, D.K. (1979). Usable knowledge. New Haven: Yale University Press.

Marris, P., & Rein, M. (1973). Dilemmas of social reform. Chicago: Aldine.

Morris, L.L., & Fitz-Gibbon, C.T. (1978). Evaluator's handbook. Beverly Hills: Sage.

Newman, D.L., Brown, R.D., & Braskamp, L.A. (1980). Communication theory and the utilization of evaluation. In L.A. Braskamp & R.D. Brown (Eds.), New directions for program evaluation. San Francisco: Jossey-Bass.

Partlett, M., & Hamilton, D. (1976). Evaluation as illumination: A new approach to the study of innovatory programs. In G.V. Glass (Ed.), Evaluation studies: Annual review (Vol. 1). Beverly Hills: Sage.

Patrick, E., McCann, R., & Whitney, D. (1981). The dissemination linking process: A view from the regional exchange. Washington, DC: National Institute of Education. (ERIC Document Reproduction Service No. ED 206 662).

Patton, M.Q. (1980). Qualitative evaluation methods. Beverly Hills: Sage.

Patton, M.Q. (1978). Utilization-focused evaluation. Beverly Hills: Sage.

Pelto, P.J., & Pelto, G.H. (1978). Anthropological research. Cambridge: Cambridge University Press.

Powdermaker, H. (1966). Stranger and friend. New York: W.W. Norton.

Sarason, S.B. (1982). The culture of the school and the problem of change (2nd ed.). Boston: Allyn and Bacon.

Schultze, C.L. (1968). The politics and economics of public spending. Washington, DC: The Brookings Institution.

Scriven, M. (1967). The methodology of evaluation. In R.W. Tyler (Ed.), Perspectives of curriculum evaluation. Chicago: Rand McNally.

Stake, R.E. (1973). Evaluation design, instrumentation, data collection, and analysis of data. In B.R. Worthen & J.R. Sanders (Eds.), Educational evaluation: Theory and practice. Worthington, OH: Charles A. Jones.

Stake, R.E. (1977). Responsive evaluation. In R.W. Tyler (Ed.), Beyond the numbers game. Berkeley: McCutchan.

Stake, R.E. (1967). Toward a technology for the evaluation of educational programs. In R.W. Tyler (Ed.), Perspectives of curriculum evaluation. Chicago: Rand McNally.

Thompson, B., & King, J.A. (1981, April). Evaluation utilization: A literature review and research agenda. Washington, DC: National Institute of Education. (ERIC Document Reproduction Service No. ED 199 271).

Tyler, R.W. (1967). Changing concepts of educational evaluation. In R.W. Tyler (Ed.), Perspectives of curriculum evaluation. Chicago: Rand McNally.

Weiss, C.H. (1972). Evaluation research. Englewood Cliffs, NJ: Prentice-Hall.

Weiss, C.H. (1980). Social science research and decision-making. New York: Columbia University Press.

Weiss, C.H., & Bucuvalas, M.J. (1977). The challenge of social research to decision making. In C.H. Weiss (Ed.), Using social research in public policy making. Lexington, MA: Lexington Books.

Williams, R.C. & Bank, A. (1981). Use of data to improve instruction in local school districts: Problems and possibilities. In C.B. Aslanian (Ed.), Improving educational evaluation methods: Impact on policy. Beverly Hills: Sage.

Wise, A.E. (1979). Legislated learning: the bureaucratization of the American classroom. Berkeley: University of California Press.

Wise, R.I. (1980). The evaluator as educator. In L.A. Braskamp & R.D. Brown (Eds.), New directions for program evaluation. San Francisco: Jossey-Bass.

Womack, T.A. (1967). Educational evaluation: Administrative function. In W.H. Strevell (Ed.), Rational of education evaluation. Washington, D.C: Office of Education, Bureau of Elementary and Secondary Education (DHEW). (ERIC Document Reproduction Service No. ED 034 292).

Young, C.J., & Comtois, J. (1979). Increasing congressional utilization of evaluation. In F.M. Zweig (Ed.), Evaluation in legislation. Beverly Hills: Sage.